

CiteULike y Connotea: herramientas 2.0 para el descubrimiento de la información científica

Por Julio Alonso-Arévalo, José A. Cerdón-García y Helena Martín-Rodero

Resumen: Los gestores de referencias sociales automatizan una tarea repetitiva y tediosa como es la gestión bibliográfica, y ofrecen una alternativa a los motores de búsqueda y bases de datos tradicionales en favor de la mediación social y el descubrimiento científico. Se hace una reflexión sobre las implicaciones del etiquetado social en los procesos de gestión de información bibliográfica personal en el entorno 2.0 y se analizan dos de las aplicaciones más populares, aunque todavía poco conocidas y utilizadas en España: CiteULike y Connotea.

Palabras clave: Gestores de referencias sociales, Referencias bibliográficas, Web 2.0 Investigación 2.0, Descubrimiento de la información.

Title: CiteULike and Connotea: 2.0 tool for discovery of scientific information

Abstract: Social reference managers automate repetitive and tedious tasks such as literature management, offering an alternative to search engines and traditional databases for social mediation and scientific discovery. In this study we reflect upon the implications of social tagging processes for personal bibliographic management in the 2.0 environment, and we study two of the most famous applications, although still little known and employed in Spain: CiteULike y Connotea.

Keywords: Social reference managers, References, Web 2.0 Research 2.0, Information discovery.

Alonso-Arévalo, Julio; Cerdón-García, José A.; Martín-Rodero, Helena. "CiteULike y Connotea: herramientas 2.0 para el descubrimiento de la información científica". *El profesional de la información*, 2010, enero-febrero, v. 19, n. 1, pp. 86-93.

DOI: 10.3145/epi.2010.ene.12



Julio Alonso-Arévalo es el responsable de la biblioteca de Traducción y Documentación de la Universidad de Salamanca. Forma parte de dos proyectos del Grupo RCLIS: DoIS y E-LIS. Es el coordinador de la lista de distribución de biblioteconomía y documentación InfoDoc y es autor de diferentes artículos en revistas especializadas sobre acceso abierto y sobre gestores de referencias.



José A. Cerdón-García, profesor titular de la Univ. de Salamanca desde 1987, es director del Master de edición de esa universidad, y director de la revista Pliegos de Yuste: revista de pensamiento y cultura europeos. Investigadora sobre la industria editorial y las fuentes de información, y es autor de varias monografías y numerosos artículos. Preside la Asoc. Española de Bibliología.



Helena Martín-Rodero es licenciada en filología románica –especialidad en filología francesa– y licenciada en filología árabe por la Universidad de Salamanca. Es directora de la biblioteca de la Facultad de Medicina de la misma universidad desde 1990. Ha participado en diversos programas de doctorado, cursos, y en proyectos financiados por el Mº de Educación y Ciencia y la Fecyt.

EL ETIQUETADO SOCIAL, consistente en compartir palabras clave definidas por los usuarios de un sistema de información, se está planteando como un método popular para la organización de la información en internet.

Casi todos conocemos servicios de la web 2.0 que permiten a los usuarios etiquetar contenidos, como Flickr, Delicious, Tagzania, y otros. El resultado colectivo de este proceso es un complejo tejido

de comentarios, recursos y etiquetas que elaboran los usuarios en la utilización cotidiana de fuentes de información de todo tipo. Su denominación, *folksonomía* o clasificación del pueblo, es ilustrativa de su naturaleza. Aunque menos conocidas, existen aplicaciones de gestión de información científica basadas en el filtrado colaborativo que sirven para compartir referencias científicas –en lugar de enlaces a favoritos–, referencias que se con-

servan como propias, pero que son visibles para todos, y que se presentan como una alternativa muy atractiva para la minería de datos y el descubrimiento científico.

Los gestores de referencias sociales funcionan como cualquier otro sistema de indización social. Se trata de una taxonomía generada por el usuario a medida que va creando una base de datos de las referencias de sus documentos que al ser compartidas con otros, le va a

permitir navegar libremente por el conjunto de etiquetas y de esta manera conocer y compilar referencias de los documentos incorporados por otros usuarios; esto favorece los contactos con otros científicos y la formación de grupos de investigación sobre temas de interés común.

“El etiquetado social se está planteando como un método popular para la organización de la información en internet

reducen, en lugar de aumentar, la eficiencia de la búsqueda debido a la cantidad de sinónimos, homónimos, polisemia, y por la heterogeneidad y el ruido que introducen (Zanardi; Capra, 2008).

“En un área especializada las etiquetas tienen más valor intrínseco que en un contexto general, donde el término aislado podría tener diferentes interpretaciones”

En los sistemas tradicionales es el indizador, un individuo experto en las normas y formas de organización de información, el que asigna descriptores para facilitar el acceso a la información. En algunas ocasiones la indización la realizan los propios autores de los artículos a quienes la revista en la que publican exige la asignación de palabras clave a sus trabajos. En contraposición a esto, en los sistemas de etiquetado social como CiteULike y Connotea, son los propios usuarios quienes asignan estas etiquetas. Se han realizado algunos estudios comparativos que arrojan diferencias importantes entre los sistemas de indexación de usuario y los sistemas de indización profesional, pero estas cuestiones deben ser consideradas en función de la utilidad de los sistemas de etiquetado para el descubrimiento de recursos, ya que el mundo de las folksonomías incluye términos que nunca aparecerían en una indización realizada por expertos o en un tesauro, pues corresponden a un vocabulario real que utilizan los usuarios de un ámbito concreto o determinado del conocimiento (Kipp, 2009). Estos sistemas también han sido cuestionados por el hecho de que las etiquetas, al ser definidas por la gente de manera informal, cambian continuamente y, en muchos casos

El etiquetado social es el ejemplo más popular de cómo el software social ha ayudado a superar los límites de los enfoques tradicionales de la categorización de contenidos. Este conjunto de datos a gran escala formado por las mencionadas estructuras conceptuales llamadas folksonomías ha llevado a algunos autores a formular lo que han denominado FolkRank (Hotho et al., 2006), es decir, un algoritmo basado en la popularidad que genera la estructura de la folksonomía en relación a la recuperación, uso y descubrimiento de comunidades científicas dentro del etiquetado. Se podría decir que en un área especializada las etiquetas tienen más valor intrínseco que en un contexto general, donde el término aislado podría tener diferentes interpretaciones porque tiene un valor añadido, que es la especificidad de ese término en ese campo junto a otras etiquetas de su propio entorno científico. Estos datos aislados tienen un interés relativo, importante para cada investigador, sin embargo a escala colectiva suponen una solución interesante para la evaluación de los contenidos científicos a gran escala. Si un rango de etiquetas lo agrupamos por categorías científicas tendremos los llamados “índices de populari-

dad”. Estos pueden ser de utilidad para marcar tendencias en determinados ámbitos científicos en los que las referencias de los artículos más populares se ordenan por el número de veces que han sido compartidas por el resto de investigadores (Taraborelli, 2008). Parece razonable que si se está trabajando sobre un tema del que existe un documento compartido por centenares de investigadores especialistas en ese tema, ese documento será de obligado conocimiento; de cualquier modo, si no se considera una medida de calidad, sí lo es de popularidad.

A pesar de que CiteULike tiene ahora más de 3 millones de referencias e incorpora al día unas 5.000 nuevas, en España apenas ha tenido eco en la literatura profesional. Sólo se pueden mencionar uno o dos trabajos, destacando la ponencia presentada en las *Jornadas Españolas de Documentación* de 2007 por Jesús Tramullas y Mónica Giménez-López (2007) en la que se hace una evaluación del software libre para la gestión de referencias.

“Las folksonomías incluyen términos que nunca aparecerían en una indización profesional, pues corresponden al vocabulario real de los usuarios”

La función básica de estas herramientas es simple: cuando un investigador localiza una referencia a un documento que le interesa en la Web, pueden hacer clic en un botón que se instala en el navegador y agregarlo a su colección personal. Posteriormente sólo deberá añadir unas etiquetas que caractericen el contenido del mismo. Las etiquetas que se incluyen son términos del lenguaje natural significativas para

Register for free at <https://www.scipedia.com> to download the version without the watermark

él y al mismo tiempo para otros investigadores que trabajan en esa área de conocimiento (Alonso-Arévalo, 2009).

Seguidamente vamos a describir *CiteULike* y *Connotea*, pero existen otras aplicaciones que comparten características similares como son *Zotero* o *Mendeley*, e incluso las áreas compartidas de otros gestores clásicos como *RefWorks* o *EndNote Web*, aunque no son propiamente gestores de referencias sociales como lo son *CiteULike* y *Connotea*. En esta categoría también podría entrar *Bibsonomy*.

<http://www.bibsonomy.org/>

CiteULike y Connotea

CiteULike es un servicio gratuito en línea para organizar publicaciones científicas. Nació en 2004 en la *University of Manchester* y en la actualidad está patrocinado por *SpringerVerlag*. No es una aplicación de código abierto aunque su uso es libre previo registro. *Connotea* nace también en 2004 de los proyectos llevados a cabo por el *Nature Publishing Group*. A diferencia de *CiteULike*, *Connotea* es una herramienta *open source*, es decir, el código fuente permanece libremente expuesto para que cualquier persona que tenga conocimientos pueda mejorarla o mantenerla, lo que en cierta manera asegura su continuidad.

<http://www.citeulike.org/>

<http://www.connotea.org/>

“Hay diferencias importantes entre los sistemas de indización de usuario y los profesionales”

Ambas son aplicaciones web basadas en marcadores sociales di-



Figura 1. MyCiteULike biblioteca personal

señadas fundamentalmente para satisfacer las necesidades de los científicos y académicos (Emamy; Cameron, 2007). Una característica común es la existencia de un perfil público y uno privado; en el público aparecen las referencias de otros investigadores por las que podemos navegar, hacer listas de seguimiento e incorporarlas a nuestra biblioteca personal. *CiteULike* distingue entre *CiteULike* (perfil público) y *MyCiteULike* (perfil privado). En *Connotea* este perfil privado se denomina *My library* y el público *Community*.

Perfil personal

En *CiteULike* el perfil personal está organizado en una serie de ventanas redimensionables con la información del usuario: fotografía, nombre, correo... También da cuenta de su actividad reciente: cuáles han sido los últimos artículos que ha incorporado y cuándo, posts incluidos en el blog, actividad en *Twitter*; y lo más importante, un módulo de recomendaciones de nuevas referencias que han incorporado otros usuarios y que son semánticamente cercanas a las etiquetas que nosotros tenemos en nuestra base de datos, es decir, hace un cálculo de aquellos artículos que nos pueden interesar y que podemos incorporar fácilmente a nuestro gestor mediante cuatro opciones que sirven como sistemas de alerta para descubrir nuevos documentos de interés y mantenernos continuamente al día de una manera muy sencilla:

– *Recommendations*: etiquetas próximas a las nuestras.

– *Watchlists*: listas de seguimiento de otros autores y etiquetas suscritas por nosotros.

– *CitGesits*: artículos más populares de la última semana.

– *Neighbours*: vecinos, usuarios que utilizan etiquetas similares a las nuestras

“El etiquetado social sobrepasa los límites de los enfoques tradicionales de la categorización de contenidos”

Se trata de un módulo introducido en *Connotea* que utiliza un algoritmo que genera una ventana con registros que han utilizado las etiquetas que nosotros también utilizamos frecuentemente. Por ejemplo, si trabajamos en el tema física cuántica, nos aparecerán los nuevos registros que otros autores que trabajan en este tema han incorporado a sus bibliotecas personales; nosotros podemos incorporarlos a nuestra biblioteca simplemente pulsando en la opción *Copy*, etiquetándolos posteriormente.

Watchlists

Son listas de seguimiento de etiquetas o de investigadores con intereses cercanos a los nuestros; de esta manera podemos seguir las nuevas incorporaciones de colegas y mantenernos constantemente al día sobre lo que va publicándose en nuestro campo de interés. Para suscribirnos



Figura 2. Perfil personal

Register for free at <https://www.scipedia.com> to download the version without the watermark

a una *watchlist* pulsamos el botón *watchlist* (que nos dará a *watchlist* no seguir). El hecho de compartir *bibliotecas* con otros investigadores que trabajan en actividades cercanas favorece la cooperación y creación de redes profesionales entre científicos, así como el descubrimiento de campos interdisciplinares.

CitGeist

Es el ranking de los artículos que más veces han sido compartidos por los usuarios durante la última semana, aunque también podemos especificar si queremos ver los artículos que más veces se han compartido en el último día, en los últimos 14 días, 28 días, o en total. Los podemos igualmente incorporar a nuestro gestor personal.

Neighbours

Literalmente se traduce como vecinos. Se trata de una lista de

usuarios que utilizan etiquetas similares a las nuestras, o a ciertas *bibliotecas* no demos entrar para incorporar sus registros a la nuestra.

“FolkRank es un algoritmo similar al PageRank de Google”

Desde el perfil personal también podemos acceder a las nubes de nuestras etiquetas de autores y materias, y hacer un control de las mismas renombrando o borrando aquellas que no sean significativas. Cuando introducimos una etiqueta formada por dos términos es conveniente hacerlo con un guión, o un guión bajo. También se pueden añadir registros manualmente, introduciendo el url, un número de identificación de un documento,

como DOI, ISBN, PMID, o importándolos en formato RIS.

En *Connotea* el perfil personal, como dijimos anteriormente, se denomina *My Library*. Se puede crear una página personalizada, aunque hacerlo es un tanto complejo pues se realiza mediante un lenguaje de marcas. La biblioteca personal dispone en la parte central de las referencias bibliográficas que podemos editar para corregir, añadir datos o borrar, en caso de que ese registro ya nos haya dejado de interesar; así mismo mostrará si esa etiqueta es compartida por otros usuarios o si alguien ha añadido algún comentario respecto a ella. En el lado izquierdo muestra nuestros grupos y las etiquetas que hemos asignado. La parte derecha está dedicada a la interacción entre un perfil y otro por medio de una barra de herramientas que nos permite añadir un marcador, crear grupos, renombrar



Figura 3. Añadir a una lista de seguimiento o WatchList

Register for free at <https://www.scipedia.com> to download the version without the watermark

etiquetas, instalar el botón *Add to Connotea*, importar y exportar referencias además de proporcionar acceso a las herramientas de configuración y personalización del sistema.

En *Connotea* destacan dos herramientas similares a las que veíamos en *CiteULike*. Una es *Related tags* que nos sugiere etiquetas próximas a las que hemos introducido nosotros gracias a un algoritmo que mide la frecuencia de aparición; de esta manera podemos navegar por ellas e incorporar las referencias de otros usuarios. La otra herramienta es *Related users*, usuarios que tienen etiquetas compartidas con nosotros, y a los que nos puede interesar seguir -vendrían a ser como los vecinos en *CiteULike*-. *Connotea* igualmente permite descubrir nueva información y conocer aquellos enlaces que han sido compartidos

un mayor número de veces por los usuarios. Para ello debemos elegir la opción *Popular links*; si queremos conocer cuáles son las últimas referencias que han subido otros investigadores deberemos seleccionar la opción *Recent activity*.

“FolkRank se basa en la popularidad que genera la estructura de la folksonomía de cara a la recuperación, uso y descubrimiento de comunidades científicas dentro del etiquetado”

Añadir información

La manera más clásica de añadir información a cualquier gestor

de referencias sociales es mediante el botón que se instala en la barra de favoritos del propio navegador: *Post to CiteULike* o *Add to Connotea*, de forma que cuando estamos navegando por un sitio web y queremos compilar la información de un documento simplemente pulsamos sobre ese botón y automáticamente se abre una ventana emergente que nos pide que introduzcamos las etiquetas que van a caracterizar el contenido de ese documento. Una vez que hemos realizado este proceso cerramos la ventana y seguimos con la tarea que estábamos realizando.

Importar datos

Con *CiteULike* podemos importar datos de otras aplicaciones de gestión de referencias en formato RIS y BibTeX. Cuando lo hacemos nos solicita el nivel de priori-

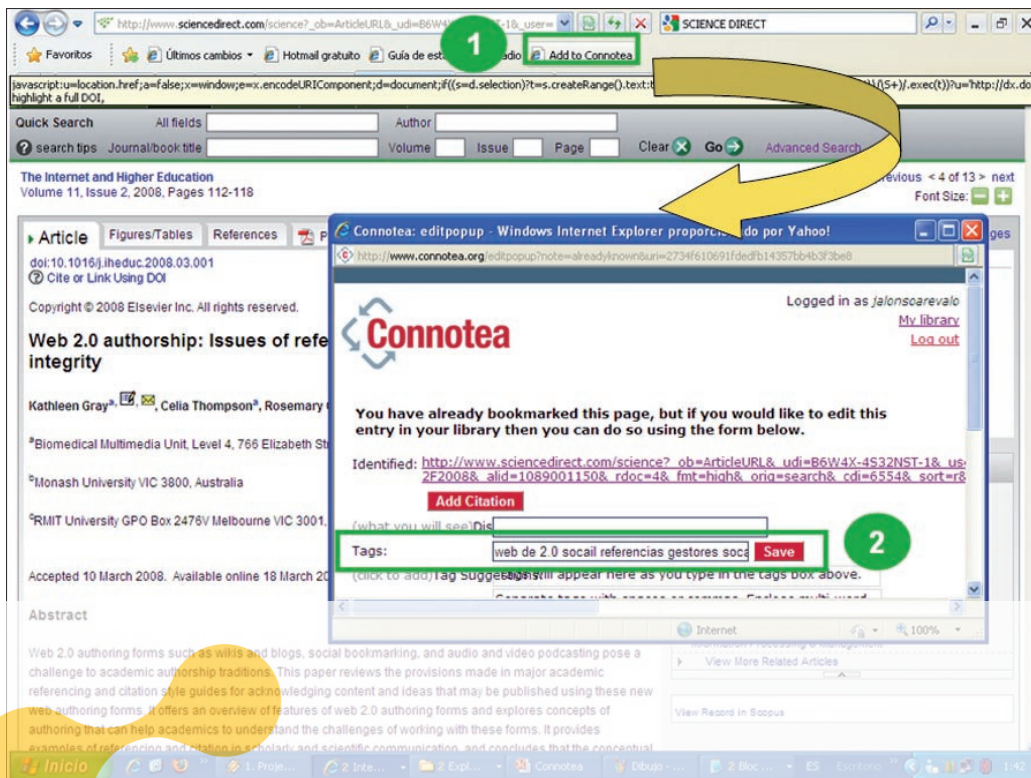


Figura 4. Botón Add to Connotea en la barra del navegador

- Añadir DOI u otra url alternativa.
- Añadir un pdf personal.
- Marcar como publicación propia.
- Establecer prioridad de lectura, o nivel de privacidad.
- Introducir una revisión del artículo.
- Incluir notas al artículo.
- Buscar artículos relacionados en CiteULike.
- Navegar por las etiquetas para buscar nuevos artículos.

- Historia del post y formatos de exportación.

dad de esos artículos, es decir, si para nosotros van a ser importantes o no; permite introducir etiquetas y establecer cómo van separadas unas de otras (espacios, comas...); igualmente ofrece la posibilidad de eliminar duplicados. Cada vez son más los sitios compatibles con CiteULike: ScienceDirect, PubMed, arXiv, etc.

de interés común. Para añadir una referencia entramos en el artículo y pulsamos sobre Copy para copiarlo en nuestro perfil privado. Estas herramientas permiten incluir un artículo externo si conocemos su número de identificación normalizado como ISBN, url o DOI (desde el perfil personal eligiendo la opción Post url en CiteULike o Add bookmark en Connotea).

“CiteGeists es el ranking de los artículos que más veces han sido compartidos por los usuarios durante la última semana”

Referencias

En cada una de las referencias, CiteULike ofrece diversas opciones para optimizar la información, su uso y el descubrimiento de nueva información:

- Exportar hasta en cinco formatos (RIS, BibTeX, pdf, rtf, Formatted Text, Delicious).
- Elegir el estilo de cita en el que queremos visualizarlo (APA, MLA, JAMA, IEEE...).
- Operar con el registro: borrar, editar, copiar, enviar al blog.
- Difundir a nuestros grupos o mediante Twitter.

Las referencias en Connotea no son tan completas como en CiteULike, aunque muchas de las tareas que realizamos desde las referencias de CiteULike es posible incluirlas en Connotea, bien editando el registro (añadir pdf, comentarios...), o bien por medio de la barra de herramientas lateral (ToolBox).

Los grupos

CiteULike permite crear grupos de investigación en torno a un tema o un departamento con la finalidad de compartir referencias entre sus miembros, y se puede hacer un blog del grupo. Los grupos son auténticas redes sociales que hacen posi-

“Las Watchlists son listas de seguimiento de etiquetas o de investigadores a las que nos podemos suscribir”

Otra forma de añadir información desde CiteULike o Connotea es desde la opción Search. Nos buscará en las carpetas de otros investigadores que hayan reseñado ese artículo, o artículos sobre un tema; de este modo podemos conocer qué investigadores trabajan en un área

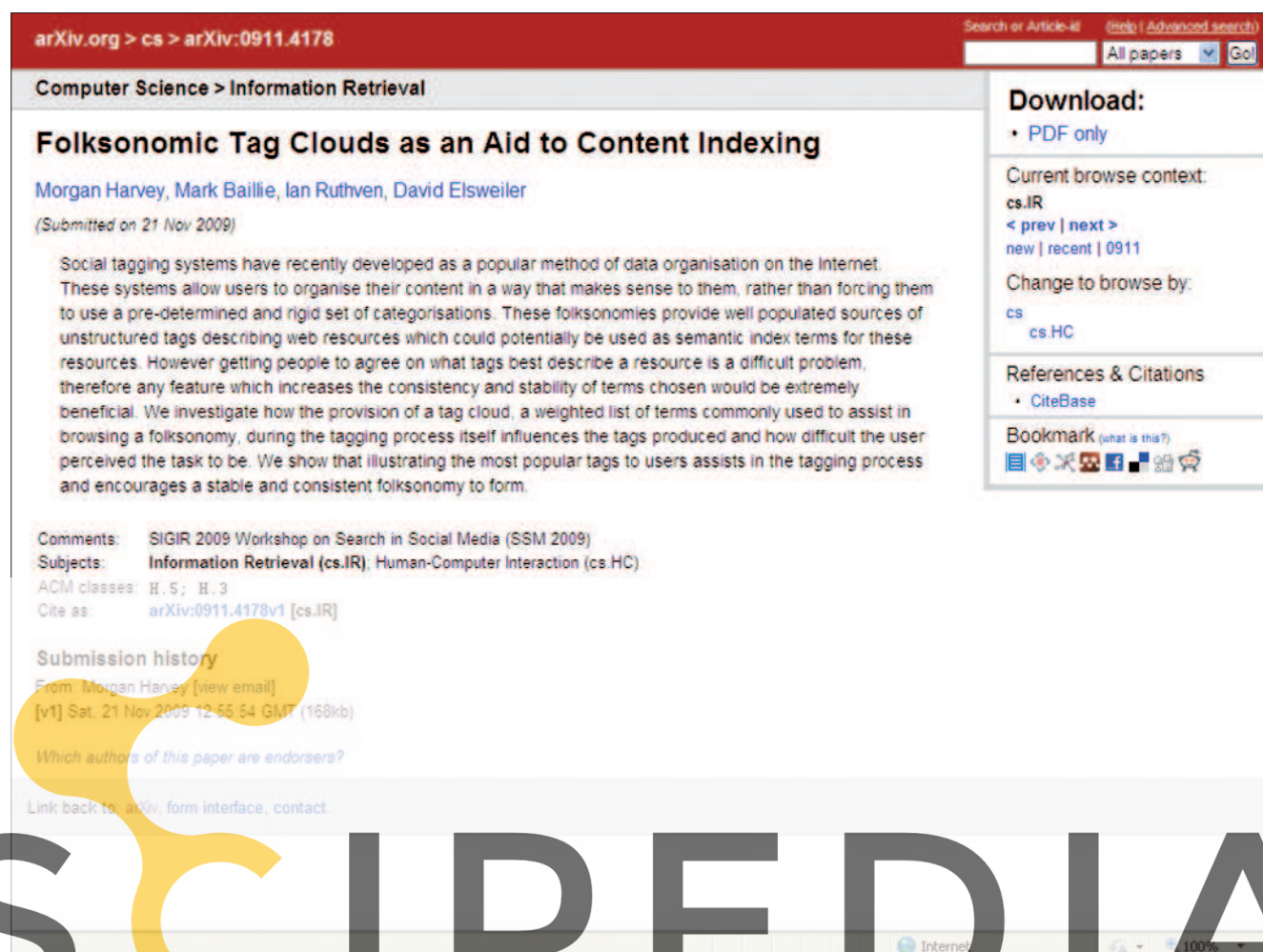


Figura 5. Fuentes compatibles con CiteULike y Connotea

Register for free at <https://www.scipedia.com> to download the version without the watermark

con el resto de los colegas de nuestro departamento, o bien con aquellos otros con los que compartimos un interés determinado.

En conclusión se trata de aplicaciones de fácil uso y comprensión, que facilitan a los investigadores la realización de un trabajo tedioso y repetitivo como es el de la compilación, almacenamiento, organización y utilización de las referencias bibliográficas; ofrecen una alternativa a los motores de búsqueda y bases de datos tradicionales en beneficio de la mediación social y el descubrimiento científico por medio de un sistema sencillo basado en el etiquetado.

Bibliografía

Alonso-Arévalo, Julio. "CiteULike o el descubrimiento de la información científica". *Comunidades de prácticas 2.0 de la Sedic*, 2009.

Alonso-Arévalo, Julio. "Connotea, software libre para la gestión de referencias sociales". *Comunidades de prácticas 2.0 de la Sedic*, 2009. http://eprints.rclis.org/17232/1/Connotea_-_SEDIC.pdf

Alonso-Arévalo, Julio. "Los gestores de referencias sociales: índices de popularidad y descubrimiento científico". *Comunidades de prácticas 2.0 de la Sedic*, 2009. http://eprints.rclis.org/17211/1/Gestores_de_Referencias_Sociales_-_SEDIC.pdf

Alonso-Arévalo, Julio. "Mendeley, el Last.fm de la investigación". *Comunidades de prácticas 2.0 de la Sedic*, 2009.

"Neighbours es una lista de usuarios que utilizan etiquetas similares y a cuyas bibliotecas podemos entrar para incorporar sus registros a la nuestra"

Alonso-Arévalo, Julio. "Zotero: la máquina de vapor del software libre para la gestión de referencias bibliográficas". *Comunidades de prácticas 2.0 de la Sedic*, 2009. http://eprints.rclis.org/17220/1/Zotero_-_SEDIC.pdf

Benvenuti, Nicola. "Social tagging e biblioteche: implicazioni e suggestioni di una 'classificazione generata dagli utenti che emerge attraverso un consenso dal basso'". *Biblioteche oggi*, 2007. http://eprints.rclis.org/12879/1/social_tagging.pdf

Capocci, Andrea; Caldarelli, Guido. "Folksonomies and clustering in the collaborative system CiteULike". *Journal of physics A: mathematical and theoretical*, 2008, v. 41, n. 22, p. 224016. <http://dx.doi.org/10.1088/1751-8113/41/22/224016>

Cordón-García, José A.; Martín-Rodero, Helena; Alonso-Arévalo, Julio. "Gestores de referencias de última generación: análisis comparativo de RefWorks, EndNote Web y Zotero". *El profesional de la información*, 2009, v. 18, n. 4, pp. 445-454. <http://eprints.rclis.org/17395/1/GestoresEPI.pdf>

Emamy, Kevin; Cameron, Richard. "CiteULike: a researcher's social bookmarking service". *Ariadne*, 2007, n. 51.

Register for free at <https://www.scipedia.com> to download the version without the watermark



Zanardi, Valentina; Capra, Licia. "Social ranking: uncovering relevant content using tag-based recommender systems. En: *RecSys'08: Proceedings of the 2008 ACM conference on recommender systems*", ACM, 2008, pp. 58-51.
<http://www.cs.ucl.ac.uk/staff/V.Zanardi/socialRank.pdf>

Julio Alonso-Arévalo, José A. Cordón-García
Facultad de Traducción y Documentación. Universidad de Salamanca
alar@usal.es
jccordon@usal.es

Helena Martín-Rodero
Facultad de Medicina. Universidad de Salamanca
helena@usal.es

¡¡ Nada más...



SCIPEDIA

Register for free at <https://www.scipedia.com> to download the version without the watermark

Mi Biblioteca

La revista del mundo bibliotecario

Suscríbete a *Mi Biblioteca* y recibirás
cada año, de manera gratuita,
el *Calendario de la Lectura* y
el *Anuario de Bibliotecas Españolas*
de la Fundación Alonso Quijano.

Tfno. 952 23 54 05
www.mibiblioteca.org